Matched Pairs Checkpoint

Question (1)

A teacher is experimenting with a new computer-based instruction and conducts a study to test its effectiveness . In which situation could the teacher use a hypothesis test for matched pairs?

A: The teacher gives each student in the class a pretest. Then she teaches a lesson using a computer program. Afterwards, she gives each student a post-test. The teacher wants to see if the difference in scores will show an improvement.

B: The teacher randomly divides the class into two groups. One of the groups receives computer-based instruction. The other group receives traditional instruction without computers. After instruction, each student takes a test and the teacher wants to compare the performance of the two groups.

C: The teacher uses a combination of traditional methods and computer-based instruction. She asks students which they liked better. She wants to determine if the majority prefer the computer-based instruction.

Feedback

A : 10

✓ Good job! Since the same group of students was measured twice (once before and then after instruction) this is a case of matched pairs design.

B : 0

✗ Incorrect: Since students were randomized to two *different* groups the data obtained from this study should be analyzed using the two (independent) samples t test. The correct answer is A.

C : 0

✗ Incorrect: The teachers collects data on the variable: "liked computer-based instruction better" (yes/no). With this data the teacher could test a hypothesis about a proportion. The correct

answer is A.

✗

Question (2)

Research suggests that identical twins have many similarities, such as body weights, even as adults. A sample of ten sets of identical twins were weighed and had their weights recorded in kilograms. The researcher wishes to test whether there is a statistically significant difference in adult weight between first-born and second-born identical twins.

Given the design of the study and the question of interest, which of the following 4 computer outputs is relevant to use?

A:

Paired T-Test and CI

```
Sample       N   Mean  StDev  SE Mean
First-Born  10  77.59  23.38     7.39
Second-Born 10  78.02  17.36     5.49
Difference  10  -0.43  11.83     3.74


95% upper bound for mean difference: 6.43
T-Test of mean difference = 0 (vs < 0): T-Value = -0.12  P-Value = 0.455
```

B:

Two-Sample T-Test and CI

```
Sample       N  Mean  StDev  SE Mean
First-Born  10  77.6   23.4      7.4
Second-Born 10  78.0   17.4      5.5


Difference = mu (First=Born) - mu (Second-Born)
Estimate for difference:  -0.43
95% upper bound for difference:  15.64
T-Test of difference = 0 (vs <): T-Value = -0.05  P-Value = 0.482  DF = 16
```

C:

Paired T-Test and CI

```
Sample       N   Mean  StDev  SE Mean
First-Born  10  77.59  23.38     7.39
Second-Born 10  78.02  17.36     5.49
Difference  10  -0.43  11.83     3.74


95% CI for mean difference: (-8.89, 8.03)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.12  P-Value = 0.911
```

D:

Two-Sample T-Test and CI

```
Sample       N  Mean  StDev  SE Mean
First-Born  10  77.6   23.4      7.4
Second-Born 10  78.0   17.4      5.5


Difference = mu (First-Born) - mu (Second-Born)
Estimate for difference:  -0.43
95% CI for difference:  (-19.95, 19.09)
T-Test of difference = 0 (vs not =): T-Value = -0.05  P-Value = 0.963  DF = 16
```

Feedback

A : 0

✗ That is not quite right. While the appropriate hypothesis test is the paired t-test, the alternative hypothesis is that there are differences, without a specification of the direction of the differences. The correct answer is C.

B : 0

✗ That is not quite correct. Since the study involved matching, through the use of identical twins, it is a matched pairs t-test, rather than a two independent samples t-test. In addition, the alternative hypothesis is that there are differences, without a specification of the direction of the differences. The correct answer is C.

C : 10

✓ Correct! Since the study involved matching, through the use of identical twins, and was interested in seeing whether there were "any differences" in mean adult weights, the correct hypothesis test is the paired t-test with a "not equal" alternative hypothesis.

D : 0

✗ That is not correct. While it is correct that the alternative hypothesis is "not equal," since the the study involved matching, through the use of identical twins, it is a matched pairs t-test, rather than a two independent samples t-test. The correct answer is C.

The following is for questions 3 and 4:

Some research suggests that first born children may have higher IQ scores than their later born siblings. Do first-born identical twins have higher IQ scores than their second-born sibling? Data from a 1998 study by Tramo, Loftus, Stukel, Weaver, and Gazzaniga were analyzed to determine whether first-born identical twins have higher IQ scores than their second-born siblings. Ten pairs of adult identical twins were assessed and their Full Scale IQ scores were calculated.

Question (3)

Let μ_1 and μ_2 represent the mean Full Scale IQ score for *all* first-born identical twins and second-born identical twins, respectively, and let μ_d be the mean of the differences in IQ score

of all identical twins (IQ score of first-born twin - IQ score of second-born twin).

which are the appropriate null and alternative hypotheses?

A:

H_0: $\mu_d = 0$

H_a: $\mu_d > 0$

B:

H_0: $\mu_d = 0$

H_a: $\mu_d < 0$

C:

H_0: $\mu_d > 0$

H_a: $\mu_d = 0$

D:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 > \mu_2$

E: Both (A) and (D) are correct.

Feedback

A : 10

✔ Correct! H_0: $\mu_d = 0$ and H_a: $\mu_d > 0$ is the correct format for the null and alternative hypothesis for this problem. Since this twin study is a matched pairs design, the hypotheses should be stated in terms of the mean of the differences μ_d . Furthermore since we are testing whether the IQ of the first born twin is ***higher*** than that of the second born twin, we would like to test H_0: $\mu_d = 0$ vs. H_a: $\mu_d > 0$

B : 0

✘ That's incorrect. It looks like there is confusion about the correct direction of the sign for the alternative hypothesis. Since we are interested in seeing whether the mean Full Scale IQ score for first-born identical twins is ***higher*** than the second-born identical twins, it would be a "greater than" alternative hypothesis. Therefore, the correct answer is A.

C : 0

✘ That's incorrect. It looks like you confused the null and alternative hypotheses. Remember, the null hypothesis is always ***no differences*** and, thus, should have an equal sign, while the alternative research hypothesis should reflect the differences that would be expected. Therefore, the correct answer is A.

D : 0

✘ That's not quite right. The correct format for the null hypothesis for the paired t-test is to use the μ_d ; therefore the, null hypothesis should be: H_0: $\mu_d = 0$. In addition, the alternative hypothesis should use μ_d and the sign should be "greater than," since we are assessing whether the mean Full Scale IQ score for

✗ the first-born identical twin is ***higher*** than the mean Full Scale IQ score for the second-born identical twin. The correct answer is A.

E : 0

✗ That's not quite right. While a) is correct, d) is not correct.

Since this is a matched pairs design, the hypotheses should be stated in terms of the mean differences μ_d . The correct answer is A.

Question (4)

The following is the (edited) output for the test:

Paired T-Test and CI

Sample	N	Mean	StDev	SE Mean
First-born	10	70.6	23.4	7.4
Second-born	10	78.0	17.4	5.5
Difference	10	-0.43	11.83	3.74

95% lower bound for mean difference: -7.29
T-Test of mean difference = 0 (vs > 0): T-Value = -0.11 P-Value = 0.544

From the output we learn that:

A: The data provide sufficient evidence to reject $H_0$and, thus, conclude that the mean Full Scale IQ score for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins.

B: The data provide sufficient evidence to reject H_0. We therefore conclude the data do *not* provide evidence to conclude that the mean Full Scale IQ scores for first-born identical twins is higher than that of second-born identical twins.

C: The data do not provide sufficient evidence to reject H_0. We therefore conclude that the mean Full Scale IQ score for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins.

D: The data do not provide sufficient evidence to reject H0. In

other words, based on the data we *cannot* conclude that the mean Full Scale IQ scores for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins.

Feedback

A : 0

✗ That is incorrect. In order to reject the null hypothesis, the P-value must be below .05. Since the P-Value is (much) greater than .05, we fail to reject the null hypothesis. The correct answer is D.

B : 0

✗ That is incorrect. In order to reject the null hypothesis, the P-value must be below .05. Since the P-Value is greater than .05, we fail to reject the null hypothesis. Also, remember that if the null hypothesis were to be rejected the right conclusion would be that the mean Full Scale IQ score for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins. The correct answer is D.

C : 0

✗ That is incorrect. While it is correct that there is not enough evidence to reject the H_0, the correct conclusion is that based on the data we cannot conclude that the mean Full Scale IQ scores for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins. The correct answer is D.

D : 10

✓ Correct! There is not enough evidence to reject the H_0; therefore, based on the data we *cannot* conclude that the mean Full Scale IQ scores for first-born identical twins is higher than the mean Full Scale IQ score for second-born identical twins.

Question (5)

In which of the following situations would it ***not*** be appropriate to use

a paired t-test to analyze the data?

A: A marriage therapist believes that many couples have different perspectives regarding the state of their marriages, prior to the start of couples therapy. In order to determine whether there is difference in marital satisfaction, prior to the start of couples therapy, each member of thirty married couples were individually given a marital satisfaction questionnaire. The mean marital satisfaction scores were then separately calculated for the group of husbands and the group of wives.

B: A psychologist was interested in determining whether fraternal twins differ in terms the degree to which they are extroverts. A sample of 50 fraternal twins was given the Eysenck personality questionnaire to assess their levels of extraversion.

C: A psychiatrist believes that a new medication (New Drug) is more effective at reducing the symptoms of depression than a placebo, sugar pill (Placebo). Sixty people with depression were randomly assigned to one of two groups: 1) New Drug or 2) Placebo and were given the assigned drug or sugar pill for 30 days. Each of the 60 participants then completed the Beck Depression Inventory and the mean depression score for the New Drug group is compared to the mean depression score for the Placebo group.

D: A researcher is interested in determining the effects of sleep deprivation on the accuracy with which mazes are completed. Thirty people are given a test of mazes on Day 1 at 9 am and again on Day 2 at 9 pm after being kept awake for 36 hours. The researcher then compares the number of errors made on Day 1 to Day 2.

Feedback

A : 0

✗ That's not correct. Since the study uses marital couples, which are a "natural pair" the appropriate statistical technique would be the paired t-test. People, who are ***randomly assigned*** to two groups are ***not matched***: the observations in one sample are not linked to another sample. Thus, the correct answer is C.

B : 0

✗ That's not correct. Since the study uses fraternal twins, which are a "natural pair" the appropriate statistical technique would be the paired t-test. People, who are ***randomly assigned*** to two groups are ***not matched***: the observations in one sample are not linked to another sample. Thus, the correct answer is C.

C : 10

✔ Correct! People, who are randomly assigned to two groups are not matched; therefore, the observations in one sample are not linked to another sample. The correct statistical technique for the example would be the two independent samples t-test.

D : 0

✘ That's not correct. Since the study uses the same participants for both a pretest measure (Day 1) and a post-test measure (Day 2), the observations in one sample are dependent on the observations in another sample; therefore, the appropriate statistical technique would be the paired t-test. People, who are ***randomly assigned*** to two groups are ***not matched***: the observations in one sample are not linked to another sample. Thus, the correct answer is C.

The following is for questions 6 and 7:

Research suggests that the pressure of being timed may interfere with performance on tests that involve mathematical problems. A fictional study was conducted with 30 6th graders. First, the 6th graders were given a math test that contained 50 problems and were told that they had only one hour to complete it (Timed Condition). The same 6th graders were later given a math test that contained 50 problems and were told that they could have as much time, as needed, to complete the test (Unlimited Time Condition). The total number of correct answers for each 6th grader was then calculated for each condition. Then, for each student, the difference between the two scores (Timed-Untimed) was calculated. The researchers hypothesized that the 6th graders would get fewer correct answers, when they took the test with a time limit, as compared to when they had unlimited time.

Tramo MJ, Loftus WC, Green RL, Stukel TA, Weaver JB, Gazzaniga MS. Brain Size, Head Size, and IQ in Monozygotic Twins. Neurology 1998; 50:1246-1252.

Question (6)

If $\mu_1$and $\mu_2$represent the number of correct answers during the Timed Condition and the Unlimited Time Condition, respectively, and let μ_d be the mean of the differences in the number of correct answers (Timed-Untimed) of all 6th graders. Which are the appropriate null and alternative hypotheses?

A:

H_0: $\mu_d = 0$

H_a: $\mu_d > 0$

B:

H_0: $\mu_d = 0$

H_a: $\mu_d < 0$

C:

H_0: $\mu_d < 0$

H_a: $\mu_d = 0$

D:

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 < 0$

E:

Both (B) and (D) are correct.

Feedback

A : 0

✗ That's incorrect. It looks like there is confusion about the correct direction of the sign for the alternative hypothesis. Recall that the for each student the difference of correct answers was computed Timed-Untimed. Since we are interested assessing whether the students would get ***fewer*** correct answers in the Timed Condition as compared to the Untimed Condition, it would be a ***"less than"*** alternative hypothesis. Therefore, the correct answer is B.

B : 10

✓ This is correct. Indeed, for each student the difference (Timed-

✓ Untimed) was computed and since we are interested assessing whether the students would get ***fewer*** correct answers in the Timed Condition as compared to the Untied Condition, the correct hypotheses in this case are: H_0: $\mu_d = 0$ and H_a: $\mu_d < 0$.

C : 0

✗ That's incorrect. It looks like you confused the null and alternative hypotheses. Remember, the null hypothesis is always ***no differences*** and, thus, should have an equal sign, while the alternative research hypothesis should reflect the differences that would be expected. Therefore, the correct answer is B.

D : 0

✗ That's not quite right. While H_0: $\mu_1 - \mu_2 = 0$ and H_a: $\mu_1 - \mu_2 < 0$ captures what we are trying to assess, since this is a matched pairs design we need to write the hypothesis in terms of μ_d. Therefore, the correct answer is B.

E : 0

✗ That's incorrect. Recall that in the matched pairs design we need to write the hypothesis in terms of the mean of the differences μ_d. Therefore the only correct answer is B.

Question (7)

The researchers analyzed the data and obtained the following output:

Summary statistics:

Sample	n	Mean	Variance	Std. Dev.	Std. Err.
Timed	50	35.3	41.438774	6.4372954	0.91037107
Untimed	50	38.7	41.030613	6.4055142	0.9058765

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
Timed - Untimed	-3.4	0.3736199	49	-9.100158	<0.0001

From the output we learn that:

A: The data provide sufficient evidence to reject H_0 . Thus, the researchers conclude that 6^{th} grade students get, on average, fewer correct answers and, thus, lower scores on math tests when taking them under timed conditions as compared to untimed conditions.

B: the data provide sufficient evidence to reject H_0. Thus, based on the data the researchers cannot conclude that mean number of correct answers on the math test for the Timed Condition is lower compared to the Untimed Condition.

C: The data do not provide sufficient evidence to reject H_0. Thus, the researchers conclude that 6^{th} grade students get, on average, fewer correct answers and, thus, lower scores, on math tests, when taking them under timed conditions as compared to untimed conditions.

D: The data do not provide sufficient evidence to reject H_0. Thus, based on the data the researchers cannot conclude that mean number of correct answers on the math test for the Timed Condition is lower compared to the Untimed Condition.

Feedback

A : 10

✔ Correct! In order to reject the null hypothesis, the P-value must be below .05. Since the P-Value is less than .05, the researchers reject the null hypothesis and conclude that 6^{th} graders get, on average, significantly ***fewer***math problems correct, when the test is timed, as compared to when the test is untimed.

B : 0

✘ That is incorrect. While it is correct that there is enough evidence to reject the H_0, the conclusion is incorrect. There is enough evidence to suggest that 6^{th}graders get, on average significantly ***fewer*** math problems correct, when the test is timed, as compared to when the test is untimed. The correct answer is A.

C : 0

✗ That is incorrect. Remember, to reject the null hypothesis, the P-value must be below .05; in this case, the P-Value is below .0001. Thus, the researchers reject the $H_0$and conclude that there is enough evidence to suggest that 6^{th}graders get, on average, significantly ***fewer*** math problems correct, when the test is timed, as compared to when the test is untimed. The correct answer is A.

D : 0

✗ That is incorrect. Remember, to reject the null hypothesis, the P-value must be below .05; in this case, the P-Value is below .0001. Thus, the researchers reject the $H_0$and conclude that there is enough evidence to suggest that 6^{th}graders get, on average, significantly ***fewer*** math problems correct, when the test is timed, as compared to when the test is untimed. The correct answer is A.

Tramo MJ, Loftus WC, Green RL, Stukel TA, Weaver JB, Gazzaniga MS. Brain Size, Head Size, and IQ in Monozygotic Twins. Neurology 1998; 50:1246-1252.